DOCUMENT RESUME

ED 216 017 TM 820 161

AUTHOR	Radwin, Eugene; And Others
TITLE	A Case Study of New York City's Citywide Reading Testing Program. National Consortium on Testing, Staff Circular No. 9.
INSTITUTION	National Consortium on Testing, Cambridge, Mass.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.
PUB DATE	Jun 81
NOTE	27p.
AVAILABLE FROM	National Consortium on Testing Project, Huron Institute, 123 Mt. Auburn St., Cambridge, MA 02138 ($2.00)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Educational Assessment; Elementary Education; Evaluation; Junior High Schools; *Program Evaluation; *Reading Achievement; *Reading Tests; Testing Problems; *Testing Programs
IDENTIFIERS	New York City Board of Education; *New York Citywide Reading Test

ABSTRACT

This summary of the history and current workings of the New York Citywide reading test program emphasizes the educational consequences of it. The testing program has become controversial since its inception as a political compromise for school district decentralization. It seems to have become overly influential because children are prepared for the reading test and coached in test-taking techniques. This preparation places undue emphasis on reading over other areas of the curriculum. The program was officially described in 1977 as serving several functions for which it was inappropriate. These were evaluating federal reimbursable programs and determining pupil class assignments. The program has since been depicted as serving other functions such as teacher evaluation and diagnosis of pupil reading skills. Major problems have been created because the program's functions are in conflict. Using the test as a means to evaluate teachers conflicts with its function of providing diagnostic information for teachers on pupil reading skills. The authors recommend focusing the Citywide Test on the function of public portrayal of the city school system, fulfilling other functions that the Citywide Test now serves, and encouraging new directions for educational assessment in city schools. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED216017

The Huron Institute
NATIONAL CONSORTIUM ON TESTING PROJECT
123 MOUNT AUBURN STREET,
CAMBRIDGE, MASSACHUSETTS 02138
617-491-5450

NATIONAL CONSORTIUM ON TESTING

STAFF CIRCULAR NO. 9

A CASE STUDY OF NEW YORK CITY'S CITYWIDE READING TESTING PROGRAM

June 1981

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

by

Eugene Radwin

with the assistance of

Grace Baisinger
Anderson J. Franklin
Wayne Jennings
Hermino Martinez

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

W. Haney

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Huron Institute
NCT Project
123 Mt. Auburn St.
Cambridge, Ma. 02138

TM 820 161

This is the ninth in a series of staff circulars prepared as part of the Huron Institute's staff work for the National Consortium on Testing. The paper is intended to spur discussion on the role of large scale testing programs and to raise ideas regarding how such programs (particularly the New York City reading testing program) can be improved. The full report on which this circular is based was supported under a grant from the Ford Foundation. Although the circular has been prepared as part of staff work for the Consortium with support from the Carnegie Corporation of New York and with cooperation of numerous educators and officials in New York City, it does not necessarily represent the views of anyone other than the authors.

FORWARD

This look into the current workings and history of New York City's Citywide Reading Testing was made possible by the assistance of many people. We would like to thank Chancellor Macchiarola for giving his office's cooperation and in particular thank the top administrators of the school system's testing efforts, Alan Blumner, Muriel Abbott, and Charlotte Frank. We would also like to thank the many district educators and school board members, school administrators and teachers, and parents who met with us for their assistance. Our thanks, too, go to Peter Purchia and Cynthia Spaulding for helping us to obtain a glimpse into what the Citywide testing means for youngsters who must take the test.

Further, we want to express our appreciation to a number of people who provided helpful comments on a draft version of this report. In addition to Drs. Blumner, Abbott and Frank, these individuals include: Michael Beck, David Cohen, George Madaus, Joseph Mehlman, Vito Perrone, Frederick Smith, and Lillian Weber.

Thanks of a different sort are also in order for colleagues at the Huron Institute -- Walt Haney, Norma Robbins, and Anne Smith -- for support in conducting and completing this project. Walt Haney, who contributed to all aspects of this work, played a particularly key role in helping prepare the more technical parts of the report. Finally, thanks are most deserving to Marjorie Martus of the Ford Foundation for underwriting our efforts, for providing guidance to them, and for forebearing our delays.

I. INTRODUCTION

One regular rite of spring for New York City school children is a norm-referenced reading achievement test. Annually, for two decades, nearly all second through ninth graders have taken a test officially dubbed the New York Reading Test but which is more frequently called the Citywide Reading Test, or simply the Citywide. The program was organized into essentially its current form with the abolishment of group IQ testing in New York City in 1964, was revamped somewhat as part of decentralization of the city's school system in 1969, and more recently has been buffeted by much controversy. This paper summarizes an extended case study of the history and current workings of this testing program. The Citywide program could be viewed from many angles, but the perspective guiding our study has been the program's educational consequences. Since this paper is a summary, it does not include footnotes or bibliography. Readers interested in such details are referred to the full report.*

II. THE 1980 CITYWIDE TESTING

What is the Citywide? One way to answer this question is simply to describe one year in the life of this testing program, namely the 1980 Citywide. On the morning of Tuesday April 29, 1980, more than one-half million NYC public school pupils in grades 2 through 9 were tested on the reading skills sections of the California Achievement Test. The testing was required under the 1969 law that decentralized the New York City school system. This state legislation requires the city's school chancellor annually

---

* The full 280 page report is available from -- NCT Project, The Huron Institute, 123 Mt. Auburn St., Cambridge, Ma. 02138.

to adminster such a test so that each school in the city can be ranked in terms of the percentage of pupils reading at or above grade level. For schools falling in the lower 45% of this ranking, community school boards are granted powers over teacher certification and selection which otherwise would remain with the central administration. In addition to this legal mandate, the 1980 testing was also motivated by a variety of other functions to which the City-wide is put, including class organization, student promotion, and evaluation of special programs.

In January 1980, the central school administration issued a contract proposal inviting potential suppliers of test instruments to bid for the 1980 Citywide. The proposal contained detailed specifications on the type of test and numbers of test booklets and other materials to be provided. Only two firms submitted bids for the 1980 Citywide contract: CTB/McGraw Hill bid its California Achievement Test (CAT), and The Psychological Corporation offered it Stanford Achievement Test (SAT). Though some officials responsible for instrument selection initially preferred the SAT, the CAT was selected--after some negotiation--mainly because it had been used in 1979, and the officials involved were concerned about continuity in the testing program.

Long before the CAT was selected, preparations for the 1980 Citywide testing were begun in schools and classrooms around the city. Test preparation activities, begun in some locales right at the start of the school year, became especially intense after January 1st. Beginning in mid-April test materials were distributed to each of 800 schools to participate in the testing. Around this time, concerns about test security arose, and in at least a few cases where 1980 test materials

had been inadvertently opened prematurely, it was arranged to use an alternate form of the CAT.

Monday, April 28, the day before the testing, was a clerical half-day designed to allow schools and teachers to prepare for the testing. On the morning of Tuesday April 29, the 1980 Citywide was administered throughout the city's elementary, intermediate and junior high schools. Following practice exercises, the test itself lasted for about an hour. Tests for grades two and three, for which children had marked answers directly in their test booklets, were sent to CTB/McGraw Hill for scoring. For higher grade levels, which used separate answer sheets, scoring was done at the central school administration. Test results were reported in a variety of ways: at individual student, class, and school levels of aggregation. As required by law, a report was later prepared which listed schools in rank order of the percentage of pupils scoring at or above grade level on the Citywide.

Initial results on Citywide testing were released to the press in September 1980. As regularly happens, the results prompted a variety of press accounts. But unlike previous years, there was also a new aspect to the 1980 Citywide testing. Through a course of events described below, an effort was made to make the 1980 Citywide results instructionally useful. This led to criterion-referenced analyses of the 1980 test results as well as the traditional norm-referenced analyses. Reports were prepared for individual students indicating mastery of the skills designed to be measured by the items on the various levels of the CAT. Such criterion-referenced analyses for individual children were distributed to teachers at the start of the 1980-81 school year. Individual reports were also prepared for children's parents.

In a formal sense, such reporting constituted the end of the 1980 Citywide testing. But many observers have suggested that Citywide test results have consequences far beyond mere reporting of results, affecting what teachers, administrators and children do in schools for years to come. To explain how this situation has come to pass, let us briefly recount the historical development of citywide testing in New York City.

III. CITYWIDE TESTING IN HISTORICAL PERSPECTIVE

The episode which best represents the initiation of citywide testing was a 1911 school survey. After several years of conflict with school officials over budgetary matters, city officials retained Paul Hanus of Harvard University to survey the efficiency of the city's school system. The survey included administration of a standardized arithmetic test to over 30,000 of the city's school children. The survey results were controversial, but nevertheless contributed to the establishment of a central school research bureau, which was to be much involved in subsequent citywide testing.

In the decade or so following its establishment in 1913, the bureau promoted the use of standardized tests in the schools, helped to develop new tests and carried out a variety of studies on special topics. One such special study, of the problem of grade retardation, recommended that children be grouped by ability for purposes of instruction and that tests be used to identify children's abilities. In the early 1920s the research bureau used standardized tests as a part of its investigation of newly established experimental schools. In the mid-1920s, the bureau began a six-year-long program of large-scale testing to gauge how well the schools

were imparting knowledge of fundamental subjects and reading, writing, arithmetic and language skills. This assessment, often referred to as "citywide" testing, was viewed with some ambivalence. On one hand it was described as not intended to make invidious comparisons among schools. But on the other, it was said to be useful for supervisors and teachers to learn "how the results attained in their school compared with those attained in other schools."

Such citywide achievement testing went into a lull in the early 1930s; when it resurfaced later in the decade, it focused more than it had earlier on reading achievement. But, as before, testing in the 1930s was embroiled in political conflict. In the early 1930s, the conflict was between the Democratic controlled city and school administrations on one hand and political reformers and Republican party politicians on the other.

The 1940s were years of growth for citywide testing. J. Wayne Wrightstone was appointed to head a new Division of Tests and Measurements in the research bureau in 1940. He remained as the school administration's chief testing expert for the next three decades. By 1949, Wrightstone had established a program of annual, citywide testing of intelligence, reading, arithmetic, and other subjects at various grades. The testing was intended to provide a picture of the effectiveness of instruction and to be useful in planning, supervising, and policymaking.

By this time intelligence testing had come to have a prominent place in citywide testing. In New York, as elsewhere, ability grouping had been introduced in the early years of the twentieth century as a means of increasing school efficiency. In the 1920s the research bureau actively engaged in promoting intelligence testing for the purpose of ability grouping.

and by the end of the 1920s this practice was reportedly widespread in New York City schools. Ability testing and grouping were connected to two other important ideas: modification of both the course of study and promotion standards according to children's abilities. Together these ideas and practices led to a policy of nearly 100% promotions by the middle of the century.

Like ability testing, diagnostic testing had been in the wind in city schools for several decades prior to 1940, but thereafter its use increased substantially. Both reading readiness and reading growth tests were used at least in part as diagnostic instruments, and diagnostic analyses of test results were also undertaken as part of efforts to meet the needs of what were then called "culturally disadvantaged" children.

Until 1963, citywide testing looked much as it had in the late 1940s. But over this period the educational context in New York City was changing markedly. Black and Puerto Rican children were coming to make up an ever increasing proportion of the school-aged population. Following the Brown decision in 1954, charges were raised that New York schools were racially segregated. Initially test results were used to highlight the needs and problems facing these minority-group youngsters. But by the 1960s, tests came to be seen not only as a means to indicate educational problems, but as part of the problem as well. Specifically, concerns arose that intelligence tests were unfair to minority children, and such concerns led in 1964 to a decision by the city's Board of Education to abolish regular group IQ testing.

As IQ testing was abolished, the number of grades and subject areas included in the program of regular, citywide achievement testing was to be greatly expanded. This was in part done so that the many functions and

influences that intelligence testing had served would not again fall on any single city test program. Within two years, however, an expanded reading achievement testing program--now annually encompassing virtually all second through ninth graders--was largely what remained of the expanded achievement testing program generally. The citywide reading achievement testing thus became, in effect, the Citywide testing program. The testing's prominence was quickly enhanced as a result of decisions by the school superintendent in 1965 to initiate a "show and tell" policy about test results and in 1966 to release to the press individual school and district Citywide results.

IV. THE CITYWIDE IN THE 1970s

The decade of the 1970s was one of turmoil for the New York City public school system in general and the Citywide in particular. In 1969 the City school system was decentralized into community school districts each with its own locally elected school board. The reorganization resulted from increasing concern over the quality of education in the City, particularly in regard to how well the school system was meeting the needs of minority children. The Bundy report which recommended the basic decentralization plan also proposed a program of biannual testing as part of a system of "checks and balances" designed to protect educational standards under a system of community control.

The decentralization legislation eventually passed by the New York legislature, however, placed an annual reading assessment in a different context. Reading testing was mandated in parts of the decentralization law concerning teacher selection and dismissal. The testing was incorporated into procedures, as we have noted, that granted local school boards more authority over teacher selection and certification for low ranking schools than they would otherwise have. Specifically, school boards could apply

these alternate selection and certification procedures in the case of schools which fell below the 45th percentile of school rankings on the Citywide. These provisions clearly were a response to minority and community groups' demands for more power over local schools and who taught in them. But the provisions also were a compromise, allowing more local control for low scoring schools, but at the same time basically affirming the established teacher selection and certification procedures favored by the school administration and the teachers' union. Despite the controversy it has spawned, the alternate teacher selection procedure appears to have had relatively little impact on the City's teaching corps, given the impacts of fiscal constraints and declining enrollments in the 1970s.

In the decade of the 1970s, four different test series were used for Citywide testing: the Comprehensive Test of Basic Skills (CTBS) and the California Achievement Test (CAT) published by CTB/McGraw Hill; and the Metropolitan Achievement Test (MAT) and the Stanford Achievement Test (SAT) published by The Psychological Corporation. The MAT was used from 1971 to 1974 and again in 1979; the SAT in 1975 and 1976; the CTBS in 1977, and the CAT in 1978 and 1979. The administration of two tests, the MAT and the CAT, in 1979 involves one of several controversies marking the Citywide during these years. These controversies, along with the yearly test results themselves, have clearly contributed to maintaining the Citywide's prominence in the 1970s.

In the 1960s and early 1970s charges of cheating on the Citywide periodically surfaced, but not until 1974-75 did this issue reach the proportions of a public scandal. Less than a week after the administration of the 1974 Citywide, television stations and newspapers reported

that children had been coached on actual items on the Citywide test prior to its administration. An investigation was launched, and it was concluded that as many as 8% of scores may have been affected by improper preparation. Later it was estimated that 2,500 of the city teacher corps may have been involved in such improprieties. Though initially the central administration had promised severe disciplinary action against anyone found to have engaged in cheating, little such action actually occurred. Instead the primary responses to the cheating issue in 1974 seemed to focus on quieting the controversy and restoring public faith in the Citywide.

In 1975 and 1976 preparations were made to insure that the cheating problem did not recur. An altogether secure test, never before available to the public, was selected for administration both of these years--specifically different forms of the SAT. Also new security precautions were introduced concerning distribution, administration, and collection of test materials. But the new procedures raised other problems. Harcourt Brace Jovanovich, the publisher involved, charged the city an additional $130,000 in 1976 for standardization and norming of the secure test, and these extra charges helped prompt an investigation by the city comptroller. The comptroller's report criticized the lack of competitive bidding for the Citywide contracts from 1964 to 1975 and, in particular, the contracting with a single publisher, Harcourt Brace, for twelve years in a row.

In response to the comptroller's report, a competitive selection process was initiated in 1977. CTB/McGraw Hill thereupon won a Citywide contract for the first time. The CTBS form selected was not a "secure" form. No such form had been sought, apparently as a result of the comptroller's charges that Citywide program was too expensive. As time for the 1977

Citywide neared, many reports surfaced of improper test preparations. The reports became sufficiently extensive that just one day prior to the testing, the chancellor ordered the Citywide postponed. The same day a small group of educators and parents obtained a temporary court order against the 1977 test's administration, charging that illegal preparation was widespread and that therefore results would be meaningless.

After a brief investigation by the central testing office, the chancellor claimed the problems were few and the testing could take place. The group of parents and educators, however, continued to press their claims. A compromise was finally worked out whereby the Citywide was administered on May 11, but the test results were not to be used until the court case was fully resolved. In July the state's highest court ruled against the plaintiffs, and strongly criticized the lower courts for interfering, maintaining that the matter should have been left to the professional judgment of school officials. In addition to the court case, the 1977 controversy prompted investigations of Citywide testing practices by the school system's Inspector General, the city comptroller, and the chairman of the State Assembly's Education Committee.

The next year, 1978, was a relatively quiet one in the life of the Citywide. But it was only an interlude. Following a selection process which now involved parents as well as educators, The Psychological Corporation's new MAT was chosen for administration in 1979. Though there was some criticism of the new MAT prior to administration, what prompted the unusual controversy in 1979 were the test results. These showed that fifth and seventh graders had virtually the same median grade

equivalent scores, 5.6 and 5.7 respectively; but that the sixth graders' median score was 6.3.

The school chancellor quickly called a halt to distribution of test results, pending an investigation of the anomalous pattern of scores. The Psychological Corporation agreed that something was amiss and soon concluded (a) that the problem derived from the equating of intermediate and advanced levels of the MAT and (b) that the only remedy required was a revised set of norms for the advanced level of the MAT, which was quickly provided. However, the school administration did not accept the firm's explanation nor its solution. (Later, independent reviews of the anomalous results by testing experts appointed by the chancellor concluded that the solution proposed by The Psychological Corporation did not necessarily solve the problem.) Hence the chancellor ordered a second Citywide test in 1979, this time using CTB/McGraw Hill's CAT. Announcement of the retest was not greeted with pleasure. One community district boycotted the retest. Despite strenuous counter-measures by the central administration, the boycott was largely successful.

The retesting in spring 1979 had ramifications for a new program of diagnostic testing planned for the following fall. The central administration had been working for some months to develop a citywide program of diagnostic reading testing, designed to provide results which would be instructionally useful. The Psychological Corporation, apparently in response to the problems which had arisen over the 1979 Citywide, chose not to submit bids that spring for the diagnostic program. (Its MAT and SAT had been approved for the bidding.) In consequence, the CAT was the by-default winner. This was ironic. The CAT had been approved to be bid

largely to ensure that the selection process was competitive. Moreover, by the time the bids were opened, the CAT had already been administered as a standard norm-referenced instrument for the 1979 retesting. Rather than readminister the CAT again in the fall, it was decided simply to apply a criterion-referenced analysis to the spring test results in order to obtain results intended to be instructionally useful.

This unanticipated merging of a diagnostic testing function onto the traditional, norm-referenced Citywide test led to several complications. First, the diagnostic results distributed in the fall 1979 were somewhat out-of-date since they were based on a test given the previous spring. Second, high rates of pupil mobility made the mere distribution of spring results difficult in the fall. Third, the change in test dates disrupted plans for Title I program evaluations. But despite these problems, the fact that the diagnostic analyses took place at all was something of an accomplishment given the turmoil over Citywide testing in 1979.

In 1980, the CAT was employed for Citywide testing. This time it was intended from the start that the one testing would receive both norm-referenced and criterion-referenced, "diagnostic" analyses. Shortly prior to the testing, responsibility for Citywide was shifted from the central evaluation office to the central curriculum office. The move was part of an effort both to make the testing more instructionally relevant and to centralize curriculum and standardize promotion practices citywide. What remains to be seen, however, is how such efforts will fit with new statewide promotion standards and the new statewide competency test. In short, just as it had at the start of the decade, Citywide testing in 1980 has become embroiled in conflict, or at least potential conflict, over who should exercise authority for governing the schools of the city.

V. THE CITYWIDE IN OPERATION: A GROUND UP VIEW

To delve into how the Citywide works in practice, we interviewed a variety of administrators, teachers and parents--people who work, so to speak, at the ground level of New York City's public education system. Altogether some 65 people were interviewed from ten of the city's thirty-two community school districts. In addition to the author, the interviews were conducted by Grace Baisinger, Anderson J. Franklin, Wayne Jennings, and Hermino Martinez. This panel was selected to bring a range of expertise and perspectives to bear on the study. Though interviews were only loosely structured, they were organized--and later analyzed-- primarily around the dozen or so functions which the Citywide has officially been described as serving.

The Citywide's roles in public evaluation of the schools, pupil promotion and placement decisions, and diagnosis and instruction evoked the most commentary from the people we interviewed. Many of the comments were positive. People spoke favorably of the public evaluation role of the Citywide for, among other reasons, it was seen as providing a standard for the schools, offering a means of holding educators accountable, and promoting healthy competition among the schools. Similarly, considerable support was voiced for the value of Citywide test results in making promotion and placement decisions about individual children. Homogeneous ability grouping seemed quite widespread in the schools we visited and tests were said to play an especially important role in placement when a student moved from one level of school to another; for instance from elementary to junior high school. Considerable support was also evident for diagnostic type testing; however, support for the Citywide's diagnostic

value was limited. It can be summarized with the comment of one individual who said that the diagnostic information provided by the Citywide "is not perfect, but it is a start."

Negative reactions to the Citywide's evaluation, promotion and placement, and diagnostic uses were both more numerous and wide-ranging than the positive comments. Some criticisms were essentially challenges to the functions themselves and/or to the Citywide's role in serving them. School rankings were thus criticized as fostering unhealthy and pernicious competition. Dissatisfaction was also expressed that other factors in addition to the Citywide results were not included in the annual public evaluations made of schools. Similarly, both educators and parents held that test scores did not supply a sufficient basis for holding children over in grade or grouping them into particular classes, or for guiding teacher selection procedures. Some also maintained that the content of Citywide test instruments little resembled what was taught in schools. In this vein, several teachers suggested that it was absurd to have a single testing program when schools employed many different reading programs.

Another set of criticisms pertained not to the uses to which the Citywide was put but to the type of test instruments used for Citywide testing. Thus, challenges were offered to the validity of standardized norm-referenced reading tests as measures of children's reading skills. It was suggested that identification of sound/letter correspondences, isolated word meanings, and other segmented bits of reading behavior do not add up to reading skill. Criticisms were voiced that such test instruments are biased against disadvantaged youngsters and that the grade-equivalent scores used to report Citywide scores are deceptive. Criticisms were leveled

against criterion-referenced tests along similar lines--for example, that test results which describe children's mastery of specific, discrete skills are not very helpful for purposes of instruction. Complaints were also offered on the one hand that teachers lacked the training to use test results effectively, and on the other, that the timing and manner of reporting of results made them of little instructional utility.

Among the other uses of the Citywide that received much local-level commentary was its teacher evaluation function. Several administrative personnel said--and a few teachers complained--that test results were used informally in this way; however, others suggested that neither teachers nor schools were held accountable for low scores. The amount of time and work devoted to the administrative demands of testing also evoked criticism. Local-level administrators complained often about the problem of overtesting. Others criticized the nature of the testing and suggested that the only testing programs that could have a positive impact were ones developed at the local level, with teacher participation and agreement. Some of these criticisms are of course somewhat contradictory. Thus, while the central administration was criticized as not doing enough with the results of Citywide testing, it was also criticized for trying to make too much out of Citywide testing.

One topic on which people were questioned and on which we received much commentary was test preparation and cheating. Special preparation for Citywide testing is apparently widespread. Between one-half and three-quarters of the schools reportedly purchased commercially developed test preparation materials aimed at the Citywide. In some cases, children received daily test preparation exercises. Though a few administrative

personnel suggested that children are not well enough prepared for standardized testing, far more often people suggested that there was "too much prepping for the test."

Most such preparation was aimed at developing what people called test-taking skills. The idea, as one person put it, was to make children comfortable and skillful in test taking situations so that their real reading skills could "shine through." But some forms of preparation clearly crossed the border into what would be more accurately called cheating. We were told that children were sometimes drilled on actual content of the Citywide instrument prior to its official administration. Others charged that since Citywide results are interpreted mainly in terms of those scoring at or above grade-level, preparation efforts were focused on children who were only slightly below average and ignored those with little chance of having their scores raised to grade level. A variety of other practices, ranging from ones of questionable propriety to others of clear cheating, were also described. While many of the people we talked with cited instances of cheating, its actual extent is difficult to estimate, if for no other reason than its very impropriety.

Why do such practices occur? Some said that they were a natural response to the prominence of the Citywide as the major means of public accountability of the schools and the fact that most schools showed up poorly on the testing. A variety of other motivations were also described including the desire to help children against whom the tests were thought to be biased. And one principal even wondered aloud,"Maybe there is so much cheating because there is unofficial approval."

On another level, both the prominence of the Citywide, and a variety of preparation activities for it, seemed fueled by a combination of mistrust and poor communication among the school system's major constitutencies. Some administrators, for example, expressed disbelief in the competence of teachers.

The strongest support for the testing and its prominence seemed to derive from concerns of parent groups, and educators' recognition of them. Some parents clearly mistrusted "school people" and doubted their competence. As a result, some parents clearly placed more trust in tests than in teachers. And educators, aware of such mistrust, seemed to turn to tests as a source of support in making decisions about individual children. On several occasions however, educators expressed doubts about the capability of parents in understanding test results. But however imperfectly, the Citywide clearly seemed to be serving as a vehicle of communication, bridging mistrust among a variety of groups.

VI. A PRELIMINARY LOOK AT CHILDREN'S VIEWS

So far we have recounted only adults' perceptions of and reactions to the Citywide. But the group with perhaps the most stake in the testing are the children who annually take the Citywide. Thus we undertook a small scale pilot study of children's views of the Citywide test. Eighteen children altogether were interviewed, from three schools and three grade levels, just after the administration of the 1980 Citywide.

The test was fairly prominent for the children. Most children reported some kind of special preparation prior to the Citywide: distribution of special practice materials, instruction in test-taking skills, and vocabulary words to study for the test, for example. But despite prominence and special preparations, children seemed to view the Citywide as a normal part of school life. To the extent that anxiety was associated with the test,

it seemed to derive not from the test itself, but from its perceived consequences: that a child might be left back in grade level, not gain admission to a special program, or get a "whopping" from a parent.

Children were not uncritical of the test however. Some criticized its content as too "goody-goody" and not relevant to life in New York City. The children perceived a variety of reasons for the test, ranging from the government's desire "to keep things going straight" to providing students a chance to bring up their marks. In the main, the children interviewed seemed to view the Citywide as useful for others--teachers, government, school boards and the newspapers--but not something that was of any direct educational value to themselves.

VII. ANALYSIS AND CONCLUSIONS

Citywide testing in New York City, both past and present, has been controversial. But despite much recent controversy over the Citywide, and despite many criticisms of the functions served by Citywide testing and the instruments employed in it, few people sought an end to Citywide testing. Occasionally people sought an end to specific functions such as the school rankings, but more generally they wanted the testing program to serve various functions better. What is curious about this support for the Citywide is that many of its functions could be met in other ways. For example, diagnosis of children's reading problems could be accomplished by teachers based on students' regular classwork.

The broader function which seems to explain widespread support for the Citywide testing is that it appears to be serving as a vehicle for communication among various school system constituencies. The testing seems to be providing a formal "objective" language shared by both educators and parents. As such it appears to be providing a common means

of discourse where other means are often lacking, at least in part due to a combination of suspiciousness and self-defensiveness existing between educators and parents.

In part because of this, the Citywide seems to have become overly influential. Too much is done to achieve high scores on the Citywide. In some places, instruction seems to be focused on reading to the exclusion of almost everything else. Moreover for months prior to the Citywide, children are reportedly prepared not just in the substance and skills of reading but in the specifics of test-taking. The influence of the testing is also apparent in the emotions it arouses as much in adults as in children.

The Citywide is also asked to serve some functions which are simply too important to be left to a test. The role of the Citywide in allocating authority over teacher selection simply does not make much sense. This function for Citywide testing stemmed from a political compromise over decentralization, but insofar as the Citywide serves the compromise, it helps draw attention from the issues of who should teach in the city's schools and on what basis they should be selected. Also, use of the Citywide as an instrument of a centralized school improvement effort seems questionable. In previous decades, the central administration of the New York City school system has sought to use large-scale testing programs as a means of school improvement but to little apparent effect. Thus using Citywide testing once again in the 1980s as a key instrument in a Citywide school improvement program seems to be foisting on the testing program a task to which it--or at least it alone--is ill-suited. Moreover, the role which has devolved on the Citywide as the single most important public portrayer of the quality of the City's schools seems to be a task

too important to be left to any single instrument. Finally, the role being carved out for the Citywide as key determiner of grade promotion is inappropriate. Such decisions are too important to rest on one aspect--reading test performance--of the skills, knowledge, and development children display in school. In addition, because of measurement error, any one test result is intrinsically imprecise. Then, too, defining a promotion policy in terms of grade-equivalent scores seems particularly inappropriate. Grade-equivalent scores lend themselves to easy misinterpretation and can vary arbitrarily from one test series to another.

In 1977, the Citywide was officially described as serving eight different functions, ranging from evaluating federal reimbursable programs to determining pupil class assignments. Moreover, since then the testing has been depicted as also serving other functions, such as teacher evaluation and diagnosis of pupil reading skills. The sheer number and diversity of these functions--as well as of less official ones--has created a major problem: no one testing program can successfully serve so many uses well. The program and the instrumentation are being pulled in too many directions at one time. In addition, many of the uses it seeks to serve were not even gained by design but by accident. This further encumbers the program.

In many ways the Citywide appears to be a testing program overgrown. As such, it also has been undermined. In practice several of its functions conflict with one another. Using the test as a means of evaluating teachers, for example, tends to conflict with its intended role in providing diagnostic information <u>for</u> teachers on pupil reading skills. Similar conflict arises in a matter as seemingly simple as selection of a

date for the testing. Moreover, as the testing program has become overgrown and embroiled in a variety of controversies, as energies have focused more heavily on administrative matters and on test security, efforts to analyze Citywide results in ways that might illuminate changes over time or overall areas of strengths and weaknesses appear to have diminished. In short the overgrown importance of the Citywide in the educational affairs of New York City public schools has led perversely to a situation which seems to have undermined more modest and reasonable uses to which the Citywide might be applied.

VII. RECOMMENDATIONS

What then should be done with the Citywide? Given the testing program's troubled history and the many liabilities which people see in it, elimination of the Citywide might seem a reasonable possibility. But like most of the people with whom we talked in New York, we do not seek an outright end to the Citywide. Rather, we propose adapting the program so as to end the more pernicious aspects, to pare down its role in the educational life of New York City schools, and to supplement it with other sorts of assessment information. Our specific recommendations are organized in three categories dealing with (a) focusing the Citywide on the function of public portrayal of the city school system; (b) fulfilling other functions that the Citywide now serves; and (c) encouraging new directions for educational assessment in the city schools.

We propose that Citywide testing be reorganized so that it can serve as one among several means of publicly representing the status and accomplishments of the New York City school system. As a basic step toward accurately portraying the school system, the reporting of test

results in terms of grade-equivalent scores should be eliminated. Also, we recommend that individual school rankings on the basis of test results be eliminated. More positively we propose that more detailed information on the testing program and test results be made available publicly--including information on the purposes and limitations of the testing and the range of results. Moreover, we suggest that reporting of test results be incorporated into an annual, public inventory of the school system's constituents, practices, conditions, and resources.

Regarding the design of the Citywide, we suggest that instead of annually testing all pupils in grades 2-9, the Citywide should be administered to representative samples. We also suggest that the test be reorganized along lines which parallel the City's educational organization. It is proposed that that focus of Citywide testing be on community school districts rather than on individual schools or students. In recommending organization of the testing along lines parallel to those of the system's governance structure, we also suggest that plans for recentralization of the city's school system, and testing's role in that possible recentralization be given close scrutiny. We are not arguing that a decentralized school system is necessarily superior to a centralized one. Rather, given the experience of educational reform efforts in New York City over the past several decades, we maintain simply that use of structural changes in school governance and administration as instruments of school improvement ought to be closely scrutinized and not merely implemented incrementally without public recognition and discussion.

If the Citywide is to be reorganized to provide, together with other information, a public portrayal of the status and health of the city school system, we propose further that other functions previously served

mainly by the Citywide be redistributed to other types of assessment. We maintain that promotion and placement decisions regarding individual children should not be based exclusively or even mainly on the results of any single test administration. To the extent that test results are needed, more useful candidates than the Citywide tests for providing such information are local district tests or state sponsored tests--the former because they likely can be better coordinated with local patterns of instruction, and the latter because they provide a better match with state mandated promotional requirements. More generally, however, we propose that holding students over in grade not be adopted as a means of improving learning unless, as the school chancellor once proposed, special transition programs are available to meet the needs of the low-scoring children. Also, we suggest that if tests are needed to determine eligibility for and provide evaluation of reimbursable programs, then local or state tests rather than the Citywide be employed. Similarly, we suggest that if tests are to be used for diagnostic and instructional purposes, again local testing programs (or perhaps state ones) should be employed. Furthermore, we recommend that the use of Citywide test results for purposes of teacher evaluation be eliminated. As part of a broader depiction of children's skills, we suggest too that individual pupil reading test results should be presented as part of broader reading profiles, describing things such as what and how children read and their attitudes toward reading.

Finally, and more generally, we recommend that new directions in educational assessment, evaluation and reporting be encouraged. Many of the functions served currently by the Citywide represent important

needs of both the school system and the community it serves. What is doubtful is how well the Citywide or any single assessment program can serve them. For the embattled status of the Citywide testing program in 1980 represents not a failure of educational testing, so much as simply trying to do too much with too little, to foist off on the testing a wide range of functions which cannot be well-served by any single form of education assessment. The answer to this problem is not simply a new or better test--rather it is to broaden the base of educational assessment and communication. Thus, in conclusion we recommend active exploration and development of alternative approaches to educational assessment--approaches, for example, that involve observational schedules, that integrate assessment and instruction more closely, that require the active participation of teachers and supervisors.